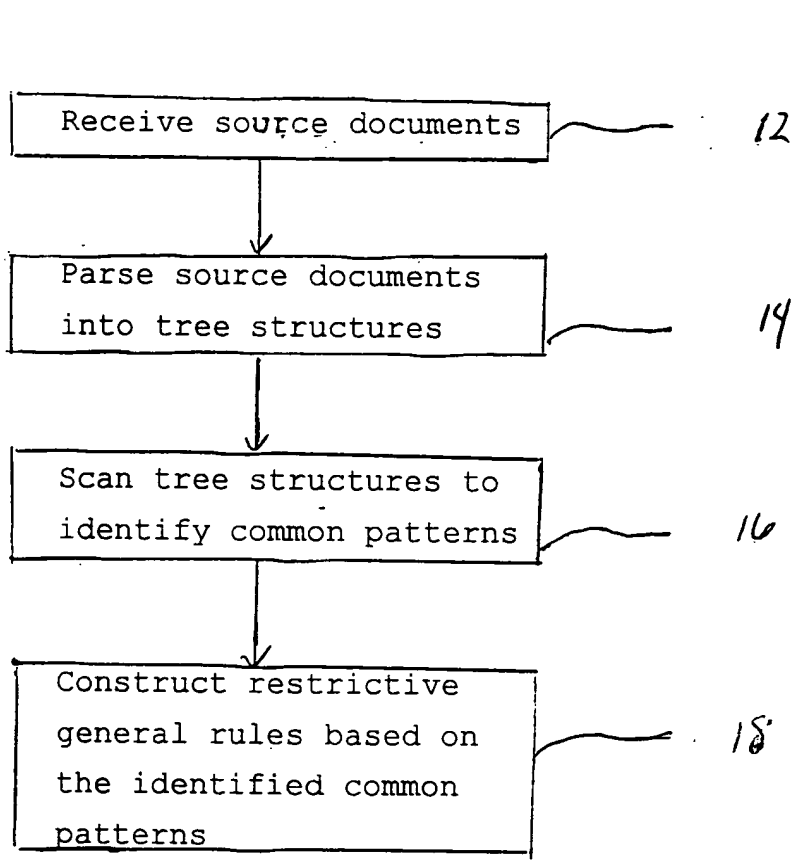

10
Receive source documents
12
Parse source documents into tree structures
14
Scan tree structures to identify common patterns
16
Construct restrictive general rules based on the identified common patterns
18

FIG. 1A

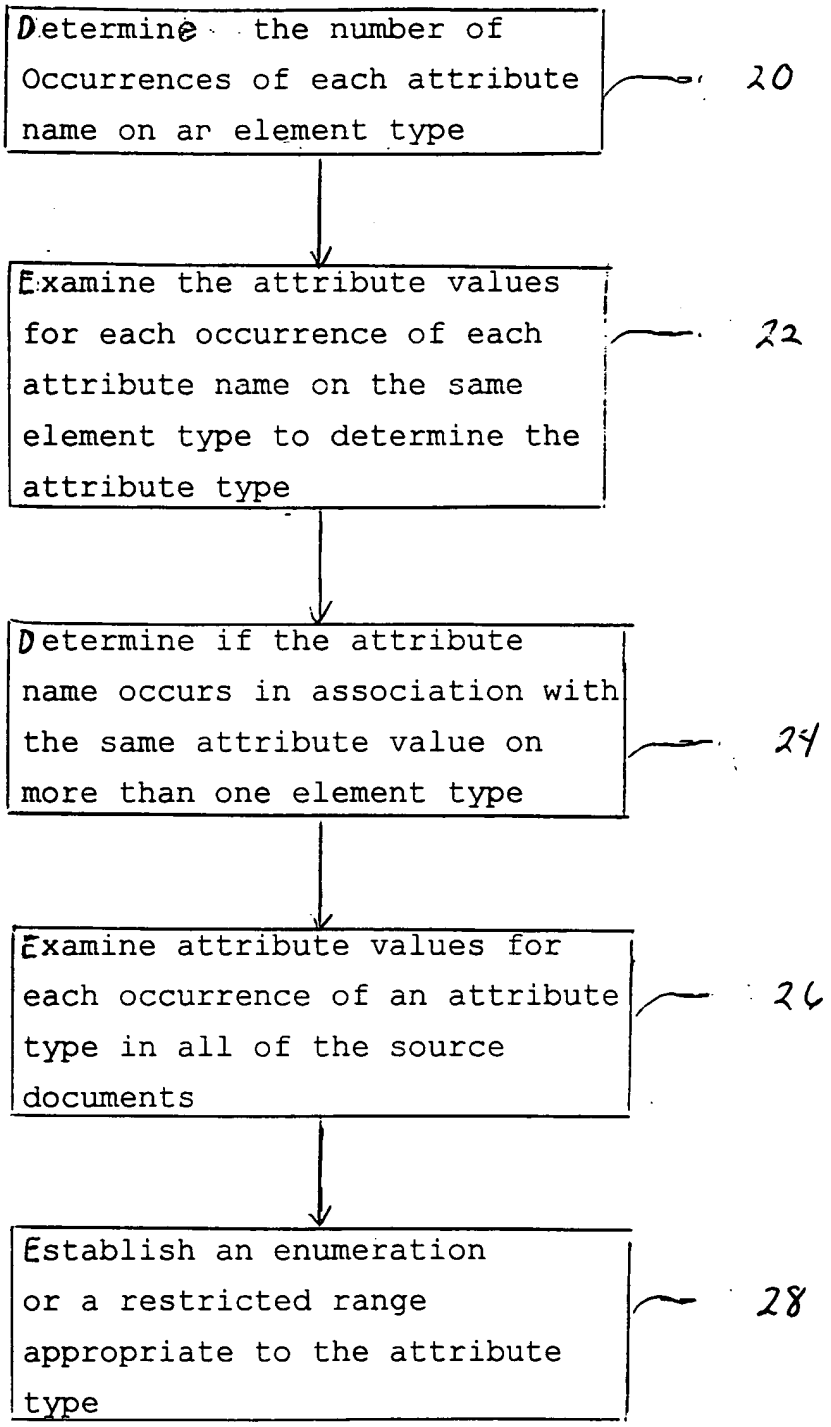

Determine the number of Occurrences of each attribute name on ar element type
20
16
Examine the attribute values for each occurrence of each attribute name on the same element type to determine the attribute type
22
Determine if the attribute name occurs in association with the same attribute value on more than one element type
24
Examine attribute values for each occurrence of an attribute type in all of the source documents
26
Establish an enumeration or a restricted range appropriate to the attribute type
28

FIG. 1B

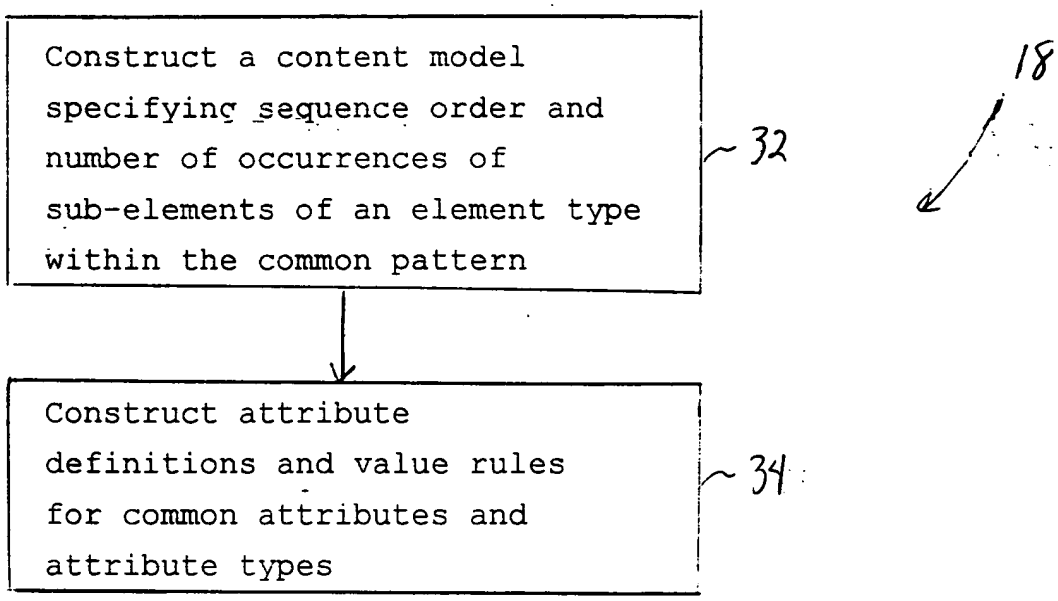

Construct a content model specifying sequence order and number of occurrences of sub-elements of an element type within the common pattern
32
18
Construct attribute definitions and value rules for common attributes and attribute types
34

FIG. 1C

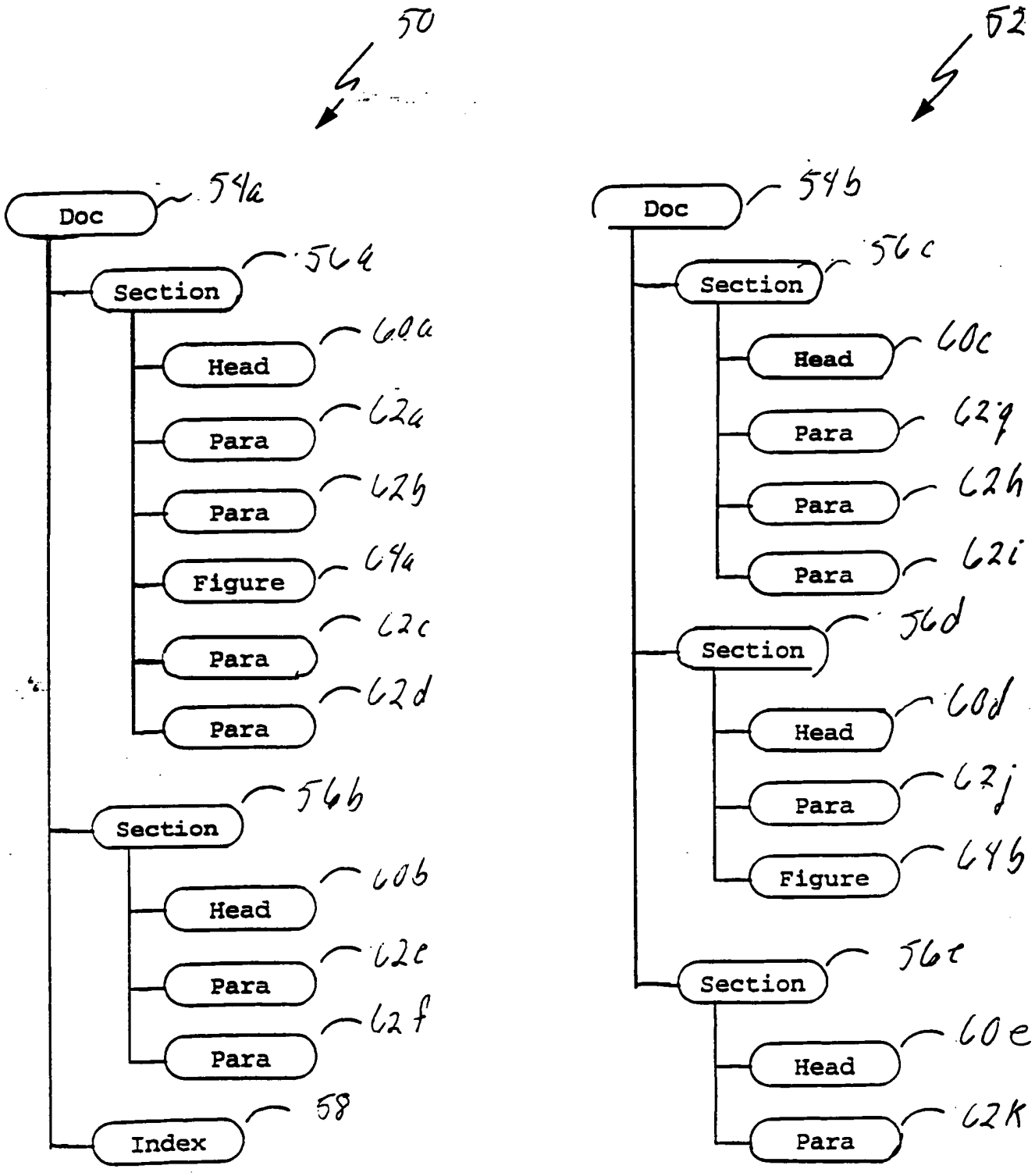

50
Doc
54a
Section
56a
Head
60a
Para
62a
Para
62b
Figure
64a
Para
62c
Para
62d
Section
56b
Head
60b
Para
62e
Para
62f
Index
58
52
Doc
54b
Section
56c
Head
60c
Para
62g
Para
62h
Para
62i
Section
56d
Head
60d
Para
62j
Figure
64b
Section
56e
Head
60e
Para
62k

FIG. 2

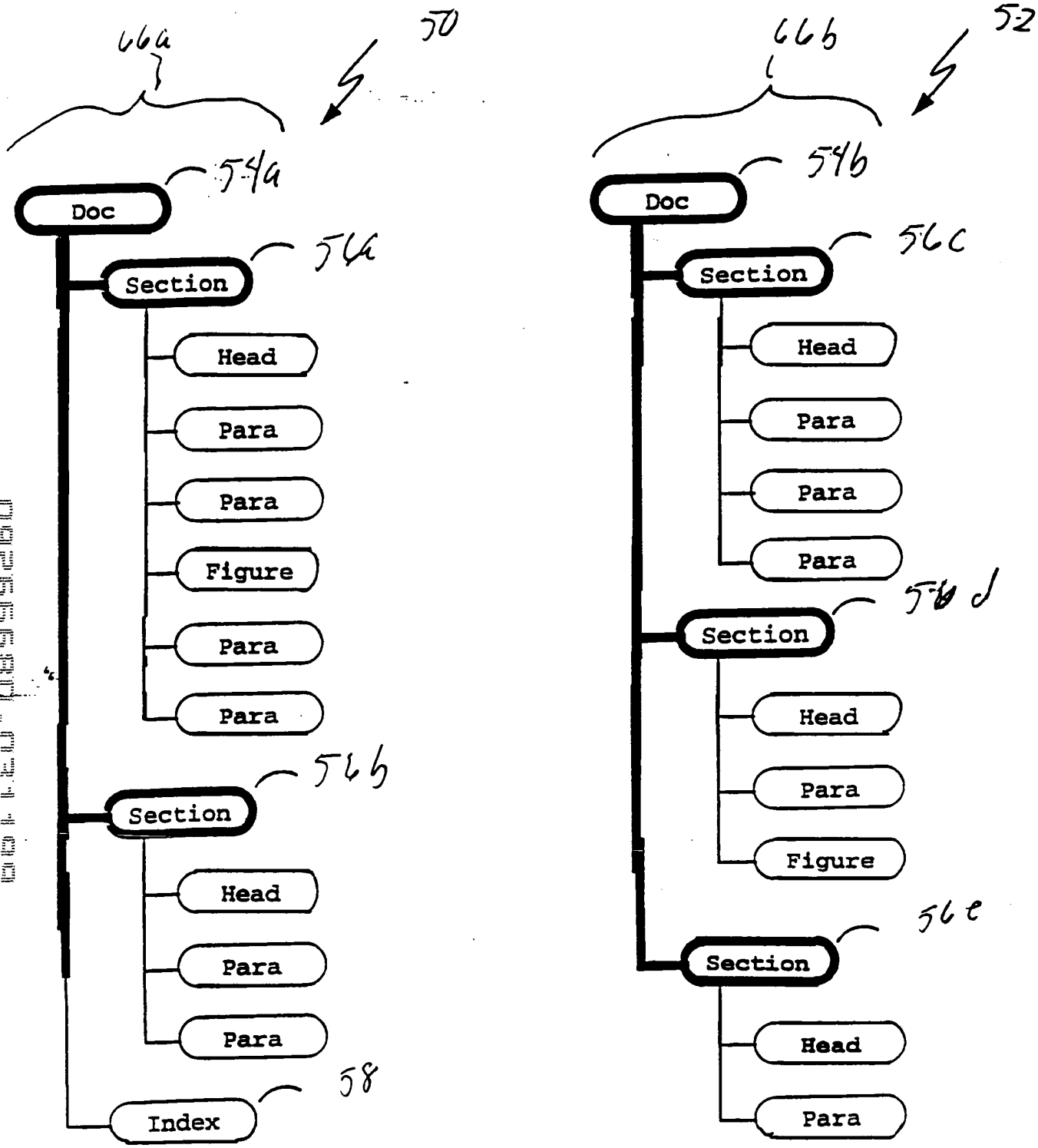

66a
50
54a
Doc
54c
Section
Head
Para
Para
Figure
Para
Para
56b
Section
Head
Para
Para
58
Index
66b
52
54b
Doc
56c
Section
Head
Para
Para
Para
56d
Section
Head
Para
Figure
56e
Section
Head
Para

FIG. 3

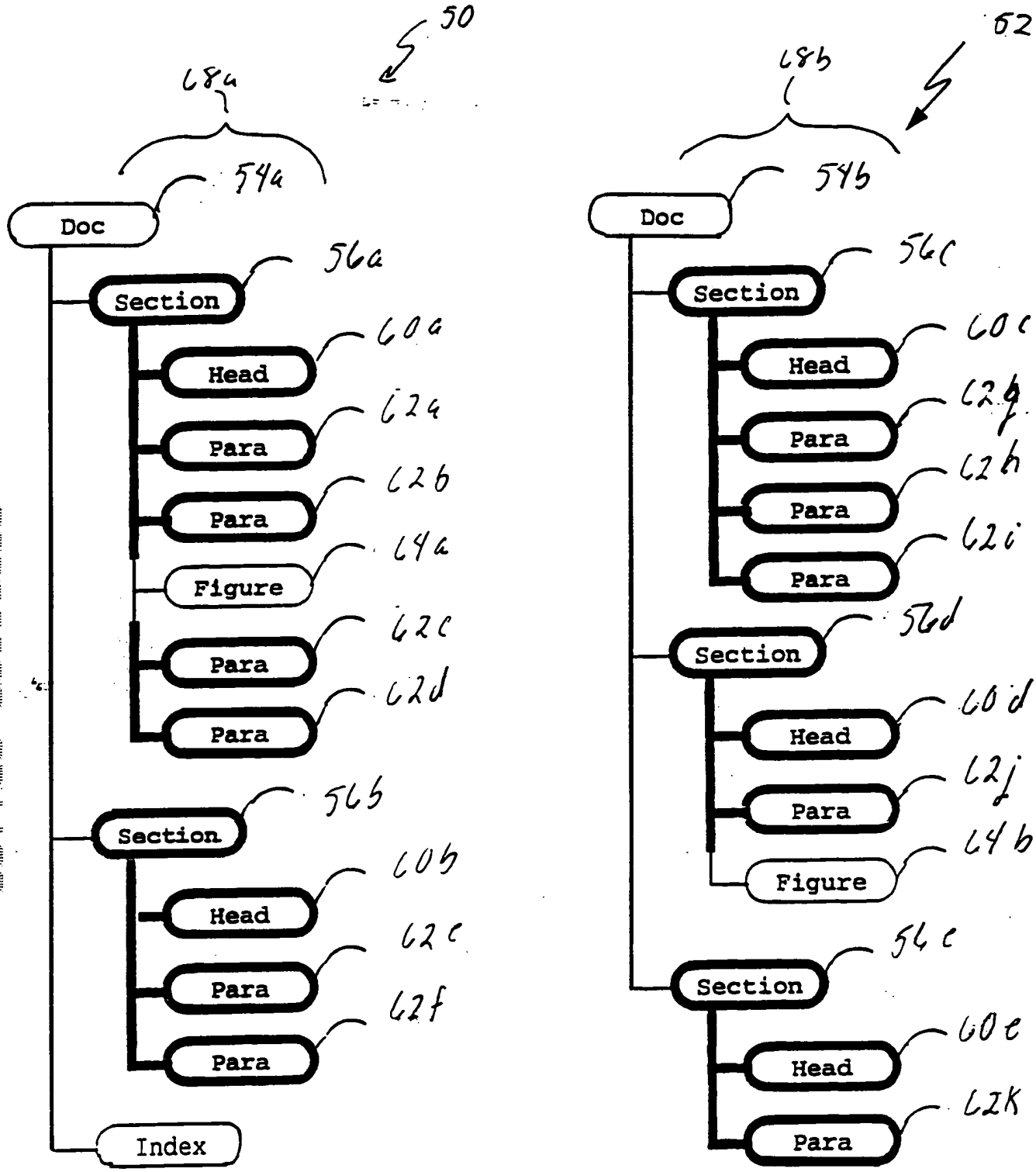

50
68a
54a
Doc
56a
Section
60a
Head
62a
Para
62b
Para
64a
Figure
62c
Para
62d
Para
56b
Section
60b
Head
62e
Para
62f
Para
Index
52
68b
54b
Doc
56c
Section
60c
Head
62g
Para
62h
Para
62i
Para
56d
Section
60d
Head
62j
Para
64b
Figure
56e
Section
60e
Head
62k
Para

FIG. 4

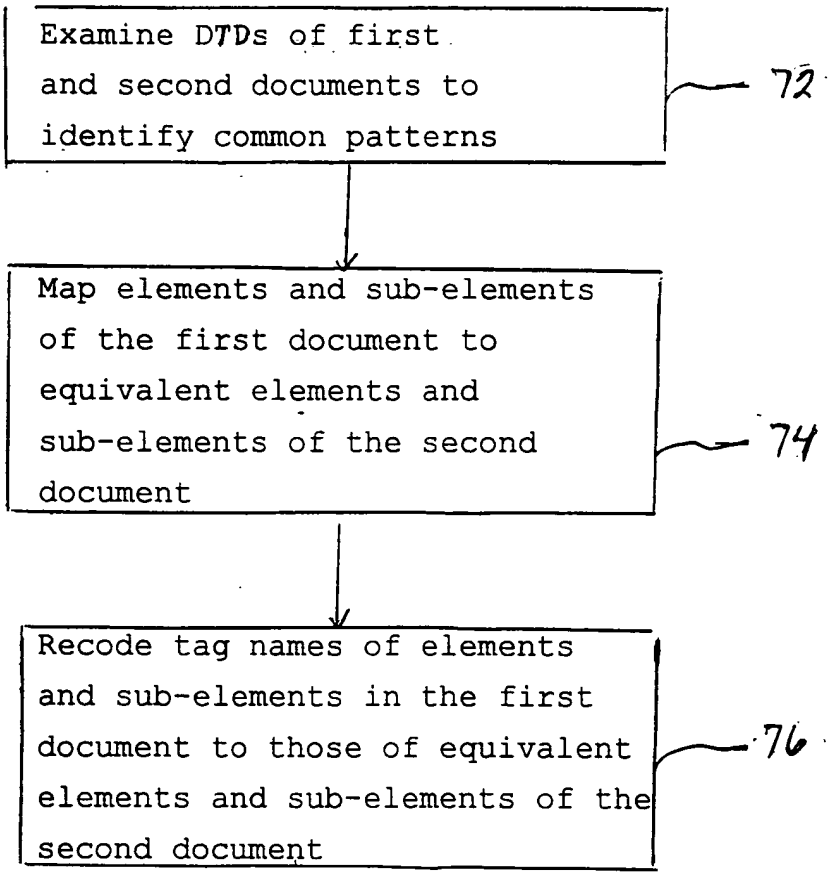

Examine DTDs of first and second documents to identify common patterns
72
70
Map elements and sub-elements of the first document to equivalent elements and sub-elements of the second document
74
Recode tag names of elements and sub-elements in the first document to those of equivalent elements and sub-elements of the second document
76

FIG. 5

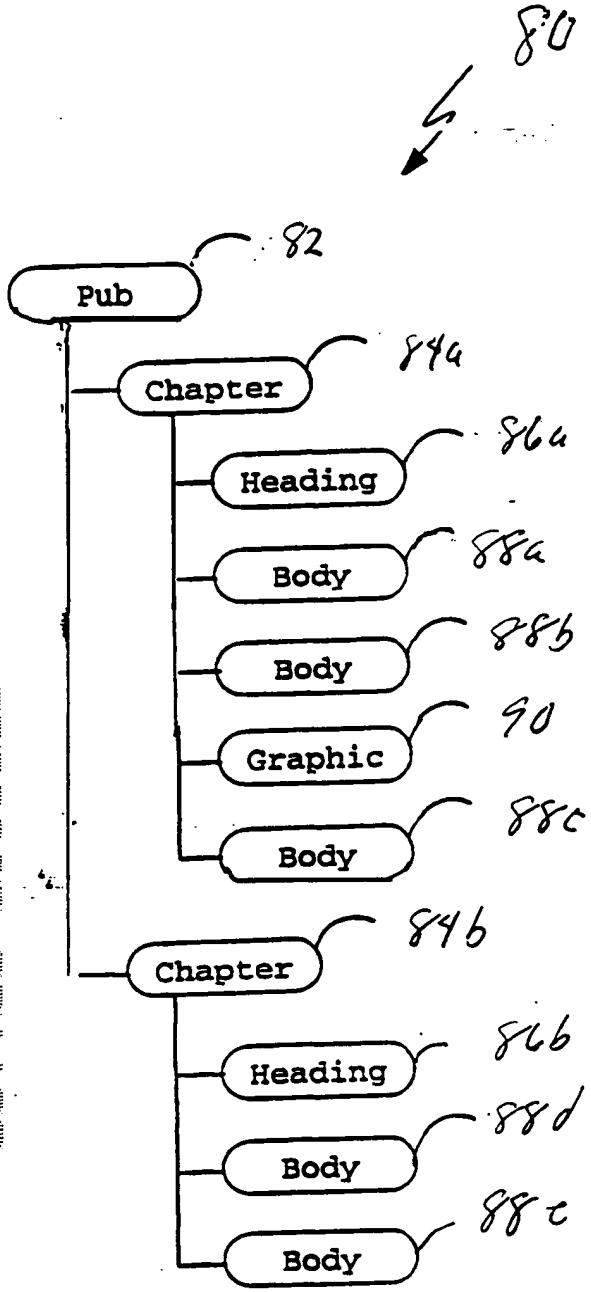

80
82
Pub
84a
Chapter
86a
Heading
88a
Body
88b
Body
90
Graphic
88c
Body
84b
Chapter
86b
Heading
88d
Body
88e
Body

FIG. 6

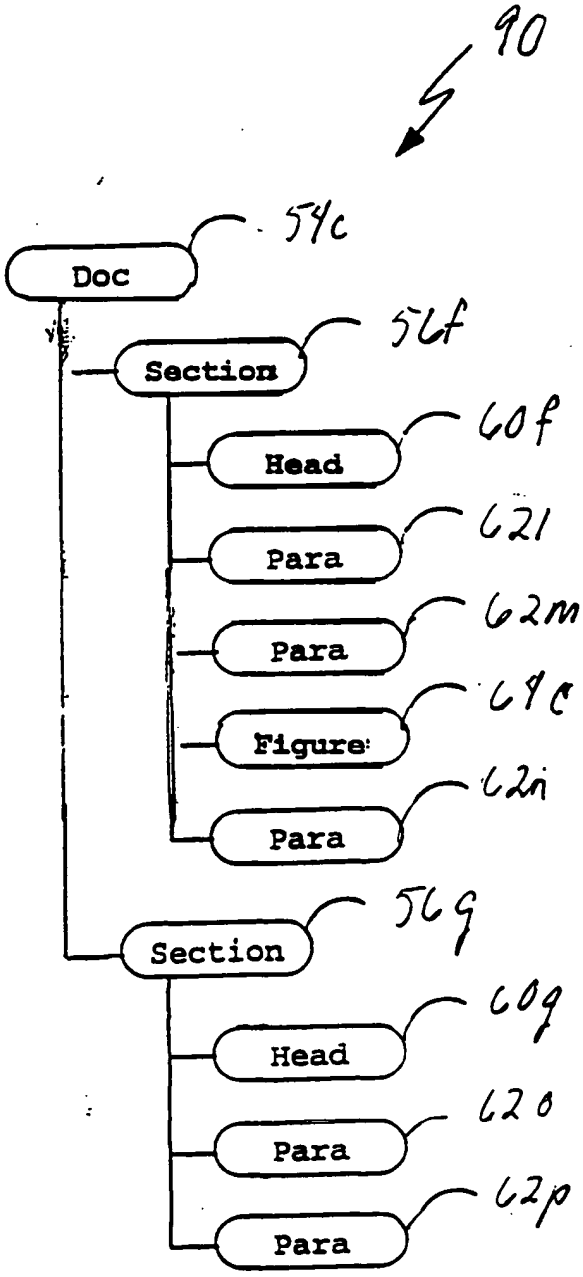
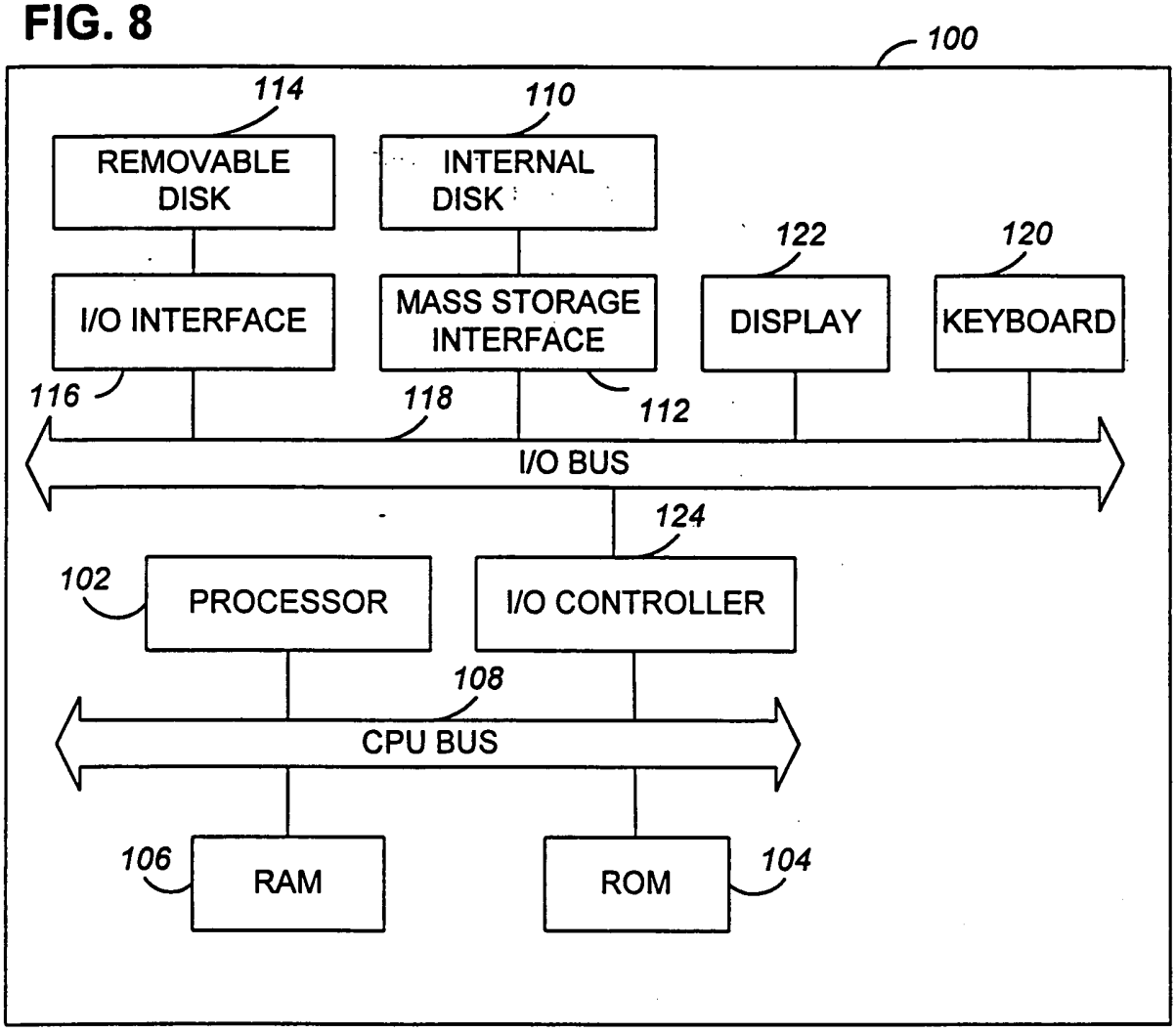

90
54c
Doc
56f
Section
60f
Head
62l
Para
62m
Para
64c
Figure
62n
Para
56g
Section
60g
Head
62o
Para
62p
Para

FIG. 7

FIG. 8


100
114
110
REMOVABLE DISK
INTERNAL DISK
122
120
I/O INTERFACE
MASS STORAGE INTERFACE
DISPLAY
KEYBOARD
116
118
112
I/O BUS
124
102
PROCESSOR
I/O CONTROLLER
108
CPU BUS
106
RAM
ROM
104